Use of suppression subtractive hybridization to examine the accessory genome of the Liverpool cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*

Catherine H. M. Smart,¹ Martin J. Walshaw,² C. Anthony Hart¹ and Craig Winstanley¹

Correspondence
Craig Winstanley
C.Winstanley@liv.ac.uk

¹Division of Medical Microbiology and Genitourinary Medicine, University of Liverpool, Liverpool L69 3GA, UK

²Regional Adult Cystic Fibrosis Unit, Cardiothoracic Centre, Liverpool L14 3PE, UK

The Liverpool epidemic strain (LES) of *Pseudomonas aeruginosa* has been highly successful at colonizing cystic fibrosis (CF) patients throughout the UK, has replaced previously established strains in CF patients, has caused infections of non-CF parents of CF patients, and can cause greater morbidity in CF than other strains of *P. aeruginosa*. Using suppression subtractive hybridization (SSH) to identify strain-specific sequences, a diagnostic test for the LES based on PCR amplification of SSH sequence PS21 had previously been developed. In this study, the SSH sequence database of LES was substantially increased, using both extension of previous sequences and new rounds of subtraction. Of 92 SSH sequences identified as present in the LES but absent from strain PAO1, 25 were assessed for prevalence amongst a strain panel consisting mainly of LES and non-LES CF isolates. Preliminary analysis of genome sequence data indicated that all SSH sequences that were LES specific or found only rarely in other strains of *P. aeruginosa* were present on one of three contigs. All of the SSH sequences screened were either unstable amongst LES isolates or were not completely LES specific. Rare false positives were found with the PS21 test. The authors suggest that a second PCR assay designed to detect SSH sequence LESF9 can be used to confirm the identity of the most prevalent CF epidemic lineage in the UK.

Received 9 December 2005
Accepted 20 February 2006

INTRODUCTION

Cystic fibrosis (CF) is the most common life-threatening inherited disease amongst Caucasians. Although patient management and infection control policies have increased life expectancy to approximately 31 years (www.cftrust.org.uk), most patients eventually succumb to morbidity and mortality due primarily to chronic bacterial infections of the lung. The most common and important CF pathogen is *Pseudomonas aeruginosa*, which establishes chronic, long-term pulmonary infections associated with a biofilm lifestyle and the production of alginate, leading to a mucoid phenotype. Once established, *P. aeruginosa* infections cannot be eradicated, and CF patients suffer decline in lung function associated with periodic exacerbations. Until a decade ago, it was generally accepted that CF patients acquired *P. aeruginosa* infections individually from the environment, and thus carried strains unrelated to those of other patients. In

1996, Cheng *et al.* (1996) reported the spread of a drug-resistant strain of *P. aeruginosa* (named the Liverpool epidemic strain, LES) amongst patients in a children's CF centre in Liverpool, UK. Subsequently other *P. aeruginosa* CF 'epidemic' strains have been reported in Australia (Armstrong *et al.*, 2003; O'Carroll *et al.*, 2004), Manchester, UK (Jones *et al.*, 2001), and the Midlands, UK (Scott & Pitt, 2004; Chambers *et al.*, 2005), raising serious questions about whether segregation strategies might be required to counter transmission. Recent analysis of post-2000 patient samples has identified the LES in 79% of 80 *P. aeruginosa*-colonized CF patients in the Liverpool adult CF centre, confirming the spread and longevity of LES infections (Panagea *et al.*, 2003). In addition to its transmissibility, the LES has been shown to replace previously established strains of *P. aeruginosa* (superinfection; McCallum *et al.*, 2001) and has infected both non-CF parents of a CF patient, causing significant morbidity (McCallum *et al.*, 2002). Furthermore, there is evidence for greater morbidity amongst CF patients colonized with the LES compared to those carrying non-epidemic strains of *P. aeruginosa* (Al

Abbreviations: CF, cystic fibrosis; FHA, filamentous haemagglutinin; LES, Liverpool epidemic strain; SSH, suppression subtractive hybridization.

Aloul *et al.*, 2004). Recently, we have demonstrated that some isolates of the LES express an unusual 'hypervirulence' phenotype characterized by early expression and overexpression of quorum-sensing-regulated virulence genes (Salunkhe *et al.*, 2005).

A survey of 31 CF centres in England and Wales, in which over 1200 isolates of *P. aeruginosa* were analysed, has identified the LES as the most common clone, being present in 48% of CF centres and accounting for 11% of the isolates (Scott & Pitt, 2004). The strain has also been identified amongst CF isolates in Scotland (Edenborough *et al.*, 2004). The antibiotic susceptibilities of LES isolates vary widely (Panagea *et al.*, 2003; Salunkhe *et al.*, 2005), and the 'gold-standard' typing method, macrorestriction analysis using PFGE, is time-consuming and requires specialist equipment. A PCR-based diagnostic test for the LES (Parsons *et al.*, 2002) has been evaluated in studies with a large number of CF isolates and found to be 100% in concordance with identification using PFGE (Panagea *et al.*, 2003; Scott & Pitt, 2004). Furthermore, the test has successfully been applied directly to patient sputum samples (Panagea *et al.*, 2003). The PCR assay is used to amplify one of a number of sequences (named PS21) identified previously using suppression subtractive hybridization (SSH), a technique that enables the identification of sequences present in one strain but absent from another (Parsons *et al.*, 2002; Winstanley, 2002). However, more recently, we and others have identified a small number of non-LES isolates that are positive using the diagnostic test (Lewis *et al.*, 2005; F. W. Scott and T. L. Pitt, unpublished data).

P. aeruginosa contains a core genome that possesses a highly conserved backbone comprising the vast majority of the genome and including most of the recognized virulence factors (Ernst *et al.*, 2003; Wolfgang *et al.*, 2003), and an accessory genome (or mobilome) that includes mobile genetic elements and genomic islands (Larbig *et al.*, 2002; Spencer *et al.*, 2003). The accessory genome of the LES is likely to contain not only strain-specific sequences that could be targeted for identification, but also novel genomic islands that may contribute to the extra abilities possessed by this strain. A genome sequence project to identify all such regions has been initiated in collaboration with the Sanger Centre, Hinxton, UK (http://www.sanger.ac.uk/Projects/P_aeruginosa/).

In this study, we report the further characterization of PS21 and two other previously identified subtracted sequences, and report the results of further SSH to identify a much larger number of sequences present in the accessory genome of LES. By using preliminary data from the genome sequence project we further group many of these sequences into islands or gene clusters. Finally, we combine the data to report the development of improved PCR assays to identify the most common epidemic strain of *P. aeruginosa* in the UK CF community.

METHODS

Bacterial strains. Isolate LES400 was obtained from a CF patient and has been used previously for SSH (Parsons *et al.*, 2002). The reference strain (driver) for SSH experiments was strain PAO1, for which the entire genome sequence is known (Stover *et al.*, 2000). For prevalence studies, we had several categories of strain panels (Table 1). Twenty-two isolates of the LES, 12 isolates of the Midlands1 epidemic strain, and 11 isolates of the Manchester epidemic strain were taken, mainly from isolates obtained during the survey of CF isolates in England and Wales (Scott & Pitt, 2004). Sixteen genetically distinct non-LES isolates from adult patients in Liverpool were included from our previous SSH study (Parsons *et al.*, 2002). In addition, we obtained 43 more recent, uncharacterized, non-LES isolates (PS21 negative) from the Liverpool adult CF unit, and 32 uncharacterized non-LES CF isolates from a neighbouring adult CF unit at Clatterbridge Hospital, Merseyside, UK. We also screened representative isolates of clone C (Römling *et al.*, 1994) and strain PA14 (Rahme *et al.*, 1995). In addition, four PS21 false-positive (non-LES) isolates were also screened (gifts of Ty Pitt and Fiona Scott, Health Protection Agency, UK).

GenomeWalker extension of sequences. Previously identified subtracted sequences [PS21, PSA and PS54 (Parsons *et al.*, 2002)] were extended using the PCR-based GenomeWalker kit (Clontech), following the manufacturer's instructions.

Construction and screening of subtraction libraries. Genomic DNA was isolated from *P. aeruginosa* LES400 and strain PAO1 as described previously (Winstanley & Hart, 2000). SSH was carried out using the CLONTECH PCR-Select Bacterial Genome Subtraction kit (Clontech) as recommended by the supplier, but with a hybridization temperature of 73 °C. In the hybridizations, DNA from the LES was used as the tester and DNA from strain PAO1 was used as the driver. PCR amplicons obtained following SSH were cloned into pGEM-T (Invitrogen). The subtraction library of *RsaI* fragments thus constructed was screened by sequencing of plasmid DNA extracted from individual clones (Lark Technologies) using M13 forward and reverse vector primers. BLAST searches at the *P. aeruginosa* PAO1 genome project web site (<http://www.pseudomonas.com>) were used to determine the presence or absence of sequences in the PAO1 genome. Tester-specific sequences were further analysed using BLASTN and BLASTX searches of the general database using the site <http://www.ncbi.nlm.nih.gov>. The relative positions of sequences in the LES genome were determined by BLASTN searching the preliminary data available at the site http://www.sanger.ac.uk/cgi-bin/blast/submitblast/p_aeruginosa. Subsequent analysis of downloaded sequence data was carried out using the program ARTEMIS.

PCR amplification screening of strains. Oligonucleotide primers (Sigma-Genosys) used in PCR assays are listed in Table 2 along with the annealing temperatures used. DNA for PCR amplification was prepared by making a suspension of a few colonies in 200 µl 5% Chelex-100 (Sigma) solution. After vigorous mixing, the suspension was boiled for 5–10 min. Following centrifugation, 150 µl of supernatant containing the DNA was removed and stored at –20 °C. Typically, 1 µl of this DNA was used directly in 25 µl volumes containing 1.25 units *Taq* DNA polymerase (Promega), 1 × *Taq*Master (Helena Biosciences), 300 nM each primer, 1 × *Taq* buffer, 2.5 mM MgCl₂ and 100 µM nucleotides (dATP, dCTP, dGTP, dTTP). Amplifications were carried out in an Eppendorf MasterCycler thermal cycler for 30 cycles consisting of 95 °C (1 min), annealing temperature (1 min) and 72 °C (2 min), with an additional extension time at 72 °C (10 min) following completion of the 30 cycles.

PFGE. PFGE of *SpeI*-digested genomic DNA was carried out using 1% (w/v) agarose gels and interpreted according to the protocol of Tenover *et al.* (1995).

Table 1. Strains used in this study

Isolates	Genotype	Source	Reference
9920, 9857, 10333, 8995, 8835, 8972, 10262, 9412, 11112, 10414	LES	CF isolates, various CF centres in England and Wales	Scott & Pitt (2004)
LES109, LES400, LES397, LES424, LES427, LESG21/23, LES416, LES417, LES430, LES431, LESB58, LESH252	LES	Known LES isolates, Liverpool CF centres	Parsons <i>et al.</i> (2002); Salunkhe <i>et al.</i> (2005); this study
9245, 8916, 9057, 9069, 9190, 9231, 9608, 9750, 9946, 10066, 10112, 10397	Midlands1 epidemic strain	CF isolates, various CF centres in England and Wales	Scott & Pitt (2004)
8799, 8877, 8715, 8763, 11057, 11171, 10633, 10845, 9081, 10234, C3373	Manchester epidemic strain	CF isolates, various CF centres in England and Wales	Scott & Pitt (2004)
73, 126, 127, 43, 103, 139, 166, 393, 176, 214, 229, 267, 289, 291, 313, 438	Non-LES, genetically different (PFGE typing)	CF isolates, Adult CF Unit, Liverpool	Parsons <i>et al.</i> (2002)
43 isolates	Non-LES (PS21 negative)	CF isolates, Adult CF Unit, Liverpool	
32 isolates	Non-LES (PS21 negative)	CF isolates, Adult CF Unit, Clatterbridge Hospital	
WM1, WM25, 043801, 044401	PS21 false negatives	Ty Pitt	
PAO1			Stover <i>et al.</i> (2000)
PA14			Rahme <i>et al.</i> (1995)
Clone C			Römling <i>et al.</i> (1994)

RESULTS AND DISCUSSION

Extending PS21 and other subtracted sequences

Using the GenomeWalker system, the PS21 sequence (previously 400 bp) was extended upstream and downstream to 1256 bp. Of the extended sequence, 750 bp shared significant similarity with the C termini of *merA* (mercuric reductase) genes/proteins in the database. The best predicted protein match was with a *Streptomyces* sp. MerA (Table 3). All of the best matches were to proteins from Gram-positive bacteria. The PS21 diagnostic sequence itself overlapped with 50 bp of the predicted *merA* gene. PS54, another sequence identified previously as unique to the LES (Parsons *et al.*, 2002), was extended from 425 to 860 bp. A predicted protein derived from nucleotide positions 131–667 contained a conserved domain found in *N*-acetyltransferases (Table 3).

FHA-like genes in *P. aeruginosa*

Although not unique to the LES, the previously identified subtracted sequence PSA shared similarity with the C-terminal sequence of *Neisseria meningitidis* FhaB, a putative adhesin with similarity to filamentous haemagglutinin (FHA) (Parsons *et al.*, 2002), and was therefore potentially virulence related. Because the PSA sequence matched strain PAO1 downstream, we extended only upstream to a total of 2751 bp. Analysis of this extended region identified sequence PSA as a replacement sequence for the C terminus of the

strain PAO1 ORF PA2462, encoding an FHA-like putative adhesion-related/haemagglutinin protein. FHA is a major virulence factor involved in the adherence of *Bordetella* species to epithelial cell surfaces (Locht *et al.*, 1993). In conducting a web-based search of the strain PAO1 genome sequence database, Croft *et al.* (2000) identified six FHA-like genes. These have been assigned ORF numbers PA0041, PA0690, PA2462, PA4082, PA4541 and PA4625 at the genome database web site (<http://www.pseudomonas.com>). There is a striking similarity between PA0041 and PA2462, which share an identical sequence over most of their length, suggesting that this pair arose as a result of gene duplication (Croft *et al.*, 2000). We carried out database searches using the C-terminal 100 amino acids from ORFPA0041 and ORF PA2462 from the genome-sequenced strains PAO1 and PA14 (available from <http://www.ncbi.nlm.nih.gov>) and found a different best match for each of these sequences and the LES PA2462 sequence, indicating that C-terminal polymorphisms in these proteins are not a rare phenomenon. Furthermore, there is a second variable region of ORF PA2462 further upstream (associated with PS9).

It has been demonstrated that the insertion of a transposon into the FHA-like genes PA0041, PA2462 and PA4625 results in reduced virulence in a rat model of chronic respiratory infection (Potvin *et al.*, 2003), suggesting a role for each of these genes in pathogenicity. By extending SSH sequence PSA, we found that the LES equivalent of PA2462 contained an alternative C terminus as well as an alternative variable region further upstream. By screening data now available

Table 2. Oligonucleotide primers used in this study

Primer	Sequence (5'–3')	Amplicon size (bp)	Target	Annealing temperature (°C)
LESA3F	TCAGCCCATAGAAATTGC	375	LESA3	58
LESA3R	TCAATCTCGTCAGCAAGC			
LESA11F	CCGAGATGAGATTGGCCA	376	LESA11	58
LESA11R	TGGTGTCTGCTGCTGTCA			
LESB7F	CTCATCAGGCTTCTTTGG	546	LESB7	56
LESB7R	TTGTTAAACGGAAGCCTG			
LESB9F	TTGAGAGCCGCATCCAGA	557	LESB9	58
LESB9R	CCGGTAGATGTCTTTCCC			
LESD11F	AACAGAGGGTGCTTGATG	622	LESD11	56
LESD11R	ATAAATCGCTGGTCATCC			
LESE4F	ACAAGAGAACACCGAGAC	367	LESE4	56
LESE4R	GAGTATTTCCCGAGCCAA			
LESE6F2	GGATATTCAAGCTGTCAG	516	LESE6	55
LESE6R2	CTGCAGACGAATCACCAC			
LESF8F	AAGAGGCATTGCAACGTC	286	LESF8	56
LESF8R	CCACCAACGTTTCCTTCTG			
LESF9F	AACACTTGCTCCATCTGC	431	LESF9	56
LESF9R	CACGATATCCAGCAAGAC			
LESG7F	AAATCCGCGAAAGTATGC	748	LESG7	58
LESG7R	CTTTCAGATTCCCGTGTG			
LESG9F	AGTTCAGGTTCTCGATCA	383	LESG9	56
LESG9R	AGGAGCCGAGCATAGTTC			
LESG11F	TGTCGGTCGAGAAAGGTC	429	LESG11	56
LESG11R	CCTCAAACCTGGATTTCTC			
LESH4F	AACTGCATAAGCTCTTCC	339	LESH4	56
LESH4R	CGACACCGAAGCCTTGAA			
LES6F	CGTTTGGCAGACCTGAAG	554	LES6	58
LES6R	AAGGTTTCCCAGGCATTG			
LES1F	TCGGCGTAATGTCCTCTA	410	LES1	56
LES1R	TGAAGCCGACGATGGAAG			
LES4F	CGCCACATGACGATAGAG	635	LES4	56
LES4R	TCTACCGTGTCGAGTCTG			
LES8F	TCAGTCGACCGGAAATGC	396	LES8	56
LES8R	CTGTCTCCTCGTCGAAGT			
LESA9F	AAAGATCAAGGAGGCAAG	348	LESA9	56
LESA9R	TCGGCTATGATGCAGTTG			
LESB3F	TGGCCGGACTTTCTATGA	268	LESB3	56
LESB3R	ATTCTCGGCCTGGCTTTC			
LESC6F	CATTAAGGTCTTGCTACG	401	LESC6	52
LESC6R	TAGATATTTTCGCTGCGGC			
LESD9F	TCAACAGAGTCCCCAACG	554	LESD9	57
LESD9R	GATGGACAGATAATGCTC			
LESE7F	TTGATGATGATGCGCGGA	274	LESE7	52
LESE7R	TCGGTCATTTTGGGCGTG			
LESF1F	CAACTACCACCGAACACG	391	LESF1	56
LESF1R	CCATCGGGATTAACCTCG			
LESH5F	TTTACGTGCGGTTGCCAG	441	LESH5	56
LESH5R	ATGCCACGTGTATATCTC			
LESH12F	GAATTGGCCATATCAGAC	785	LESH12	56
LESH12R	CGAGTCCAGAAACGCTTG			

from the LES genome sequence project, we have been able not only to confirm this variation, but also to identify downstream of the PA2462 ORF a region of repeated FHA-like sequences, including an identical truncated C terminus equivalent to the PAO1 C terminus of ORF PA2462. In reporting the genome sequence of *N. meningitidis* serogroup A strain Z2491, Parkhill *et al.* (2000) refer to repeat mediated rearrangement of the 3' ends of a gene encoding a surface-exposed FHA-like protein. The sequence downstream of this gene contained complex repeated sequences also found within the gene. Sequencing of another *N. meningitidis* strain (Z4259, serogroup C) revealed a similar repeat structure, but with additional DNA inserted into the 3' sequence of NMA0688 at one of the repeats. This suggested that the 3' end of the gene can be altered by recombination with alternative 3' ends. Furthermore Klee *et al.* (2000) have reported that the C terminus of the FHA-like protein identified as similar to the PSA sequence (*fhaB* of *N. meningitidis*) is variable between strains. The authors hypothesize that repeated sequences could provide for variation in the C-terminal portion of the FhaB protein. They also report evidence that the *fhaB* gene is present in two copies in at least some strains of *N. meningitidis*.

The PA2462 peptide sequence shares significant similarity to *Pseudomonas putida* KT2440 ORFs PP0168 and PP1449 (Nelson *et al.*, 2002). PP1449 (1508 amino acids) is the better match, and is also identified when the KT2440 genome is interrogated with PA0041 (*E* value of approximately 1×10^{-46}). PP0168 is the largest gene (8682 amino acids) in the KT2440 genome and has the most unusual sequence (Weinel *et al.*, 2002). The protein is rich in asparagine, valine and threonine, but contains no cysteine and is low in arginine. It contains two large threonine-rich repeats: the N-terminal repeat (nine units, each 100 amino acids long) and the C-terminal repeat (29 units, each 219 amino acids long). Although it has much in common with the *P. aeruginosa* genome, the only virulence-related genes that the genome of the avirulent *P. putida* KT2440 contains encode putative or confirmed adhesion proteins. Both PP0168 and PP1449 share homology with adhesin genes of enteropathogens and encode adhesins that are essential for seed colonization (Espinosa-Urgel *et al.*, 2000; Nelson *et al.*, 2002).

Thus, FHA-like proteins such as PA2462 may contribute to adhesion and colonization of disparate hosts, whether seed or human, and contain repeat sequences in the C terminus and downstream of the ORF with a probable role in mediating recombination. The biological function of such variations remains unknown, but since there may be a role for such proteins in the colonization of CF patients, further study is certainly merited.

LES subtracted library

Subtraction of the genome of strain PAO1 from that of isolate LES400 produced a library of clones, 106 of which were subjected to plasmid extraction and sequencing.

Following the removal of duplicates and non-Tester-specific sequences, the subtraction added 78 sequences to the 14 obtained in a previous study (Parsons *et al.*, 2002). The combined subtracted sequences are summarized in Table 3. Fifteen of the sequences shared similarity with mobile genetic elements, including bacteriophages D3 and D3112, both of which have been reported previously in *P. aeruginosa* (Cavenagh & Miller, 1986; Wang *et al.*, 2004). A further six subtracted sequences were associated with transposition, recombination or DNA modification. Amongst the known variable genes matching subtracted sequences were several related to siderophores. We have reported previously that the LES has a type III pyoverdine receptor (Parsons *et al.*, 2002). The SSH identified sequences related to the synthesis of both type III pyoverdine and the corresponding receptor (Smith *et al.*, 2005). In addition, the LES carries genes for serotype O6. Four of the subtracted sequences shared similarity with serotype-related genes. Whilst four subtracted sequences indicated the presence of the known *P. aeruginosa* genomic islands PAGI-1 (Liang *et al.*, 2001) and PAGI-2 (Larbig *et al.*, 2002), a further five sequences shared similarity with four linked genes from *Clostridium thermocellum* ATCC 27405, suggesting that the LES may contain a low-G+C-content genomic island related to a gene cluster in this Gram-positive bacterium (Table 3). The subtraction also identified variations in the *pilABC* cluster. Such variations between strains of *P. aeruginosa* have been reported previously (Choi *et al.*, 2002; Kus *et al.*, 2004).

As well as the gene related to mercury resistance associated with PS21, the subtraction identified a putative gene involved in silver resistance. Three subtracted sequences shared similarity with proteins that have roles in the production of polyketides (Table 3). Two of these sequences matched a regulatory protein (PltR) involved in the synthesis of pyoluteorin, which has antifungal activities (Bender *et al.*, 1999). BLASTX searches for 38 % of the LES subtracted sequences revealed either no significant matches or matches with hypothetical proteins of no known function.

Preliminary searches of the LES genome using SSH sequences

Preliminary searches of the LES genome sequence using BLASTN identified all but two of the SSH sequences. Interestingly, there was no BLASTN match in the LES genome (99.98 % theoretical coverage) for either LESC4 or LESC5, which match adjacent genes from bacteriophage D3112, suggesting that this bacteriophage may be present in isolate LES400 but missing from the genome-sequenced isolate (LESB58). Notably, many of the SSH sequences were found on common contigs containing putative genomic islands. A single contig contained, in order, the subtracted sequences LESC6, LESH12, PS54, LESF10, LESG7, LESE5, PS3, LESF1, LESB11, LESE3, LES11, LESF12, PS1, PS12, LESD8, LESB7 and LES4 (cluster 1). Further analysis indicated that sequence LESC6 is located at one edge of a putative genomic island incorporating all of the other SSH sequences with the

Table 3. Summary of LES SSH

SSH sequences sharing less than 90 % nucleotide sequence identity with the PAO1 genome were included in the table. ID, identity; NSH, no significant hits.

SSH sequence	Length (bp)	GC content (%)	Best BLASTX match/comments, with accession nos	ID (%)	Length (aa)	E value
Bacteriophage/plasmid related						
LES1	503	62.4	Phage-related integrase (<i>Pseudomonas fluorescens</i> Pf-5) (YP_260895)	50	122	2e ⁻²⁶
LES6	> 730	60.7	Gifsy-2 prophage protein (<i>Salmonella typhimurium</i>) (NP_460007)	55	245	3e ⁻⁶⁸
LESA2	764	62.3	Orf26, serotype-converting bacteriophage D3 (<i>P. aeruginosa</i>) (NP_061522)	98	254	7.3e ⁻¹³¹
LESC11	664	59.9	Orf104, serotype-converting bacteriophage D3 (<i>P. aeruginosa</i>) (NP_061582)	97	38	2e ⁻³¹
LESA7	643	58.8	Phage-related tail-fibre protein (<i>P. aeruginosa</i>) (BABO3228); probable bacteriophage protein (PA0620)	100	214	e ⁻¹³⁰
				67	203	3e ⁻⁷⁹
LESC4	324	64.8	Hypothetical protein D3112p54 (Bacteriophage D3112, <i>P. aeruginosa</i>) (NP_938261)	92	107	8e ⁻⁵⁶
LESC5	528	64.2	Hypothetical protein D3112p53 (Bacteriophage D3112, <i>P. aeruginosa</i>) (NP_938260)	93	102	2e ⁻⁴⁹
LESG3	401	56.4	Hypothetical protein F116p69, bacteriophage F116 (<i>P. aeruginosa</i>) (YP_164333)	85	34	9e ⁻¹⁰
LESE7	430	62.8	Hypothetical protein mlr8022 (<i>Mesorhizobium loti</i>) (NP_108210); may be bacteriophage related	52	50	9e ⁻⁷
LESF6	719	59.9	gp1 (Bacteriophage N15) (AAC19037)	42	163	4e ⁻²⁸
LESG5	751	57.7	gp1 (Bacteriophage N15) (AAC19037); matches same sequence as LESF6, but is different from F6	41	163	3e ⁻²⁶
LESH5	594	62.8	Hypothetical protein NCgl1675 (<i>Corynebacterium glutamicum</i>) (NP_600951); probably bacteriophage related	35	104	1e ⁻¹¹
LESB7	> 1075	52.7	RepA (Plasmid pBFp1) (AAQ94182)	68	167	2e ⁻⁶²
LESC6	> 1127	54.2	Putative plasmid recombinase (<i>Burkholderia pseudomallei</i> K96243) (YP_109853)	50	187	1e ⁻³⁵
LESC2	659	58.3	Hypothetical protein, plasmid pWW0 (<i>P. putida</i>) (NP_542801)	39	220	2e ⁻³⁶
Siderophore related						
PS5	340	55.7	FpvAIII, type III pyoverdine receptor (<i>P. aeruginosa</i>) (AAN62912)	99	113	7e ⁻⁶¹
LES8	527	61.3	<i>pvdI</i> (3), type III pyoverdine-related non-ribosomal peptide synthetase (<i>P. aeruginosa</i>) (AY765261)	99	175	5e ⁻⁹⁵
LESH4	524	62.0	<i>pvdD</i> (3), type III pyoverdine-related non-ribosomal peptide synthetase (<i>P. aeruginosa</i>) (AY765261)	92	174	6e ⁻⁸⁵
LESA1	461	55.1	Outer membrane ferric siderophore receptor (<i>P. putida</i> KT2440) (NP_743022)	57	157	1e ⁻⁴⁶
LESA3	> 584	58.6	Outer membrane ferric siderophore receptor (<i>P. fluorescens</i> Pf-5) (YP_257996)	71	196	3e ⁻⁷⁶
O serotype						
LESB12	> 596	55.4	Putative glycosyl transferase WbpT (<i>P. aeruginosa</i> O6 serotype) (AAF23993)	99	198	e ⁻¹⁰⁷
PS43	> 307	51.8	WbpO, B-band O-antigen biosynthesis (<i>P. aeruginosa</i> O6 serotype) (AAF23997)	100	102	8.2e ⁻⁵⁰
PS48	862	48.6	Wzx, O-antigen flippase/translocase (<i>P. aeruginosa</i> O6 serotype) (AAF24000)	99	287	8.9e ⁻¹⁴⁰
LESD3	592	50.8	WbpS (<i>P. aeruginosa</i> O6 serotype) (AAF24002)	55	200	4e ⁻⁵⁷
Genomic islands						
PS40	198	57.5	Hypothetical protein (<i>P. aeruginosa</i>) (AAK01512); genomic island PAGI-1	100	66	2e ⁻³¹
LESF7	> 605	62.0	Hypothetical protein (<i>P. aeruginosa</i>) (AAK01519); genomic island PAGI-1	100	148	1e ⁻⁸¹
LESG6	237	62.9	Hypothetical protein ORFC81 (<i>P. aeruginosa</i> clone C) (AAN62175); genomic island PAGI-2	100	78	1.9e ⁻³⁴
LESG7	1219	52.7	Hypothetical protein (<i>Providencia rettgeri</i>) (AAM08045)	41	124	2e ⁻²¹
			Putative excisionase ORFC56 (<i>P. aeruginosa</i> clone C) (AAN62150); ORFC56 is in genomic island PAGI-2	32	133	8.0e ⁻¹¹

Table 3. cont.

SSH sequence	Length (bp)	GC content (%)	Best BLASTX match/comments, with accession nos	ID (%)	Length (aa)	E value
Linked genes (putative island)						
LESF9	513	58.1	Type I restriction-modification system methyltransferase subunit (<i>C. thermocellum</i>) (ZP_00504950)	38	158	$9e^{-19}$
LESG2	748	57.4	Hypothetical protein (<i>C. thermocellum</i>) (ZP_00504954)	44	233	$1e^{-47}$
LESD11	1176	53.3	Possible ATPase (<i>C. thermocellum</i>) (ZP_00504957)	71	298	e^{-117}
LESB4	419	50.1	Serine/threonine protein kinase (<i>C. thermocellum</i>) (ZP_00504959); N terminus	41	39	0.031
LESE4	623	49.4	Serine/threonine protein kinase (<i>C. thermocellum</i>) (ZP_00504959); same protein as LESB4	43	205	$1e^{-38}$
Drug or metal resistance/antimicrobial synthesis						
PS21	400	42.8	NSH			
Extended PS21	1256	51.4	MerA, mercuric reductase (<i>Streptomyces</i> sp. CHR28) (AAF64138)	53	240	$5e^{-67}$
LESB9	933	62.0	Putative silver-efflux pump (<i>Dechloromonas aromatica</i> RCB) (ZP_00150991); membrane-fusion protein (<i>D. aromatica</i> RCB) (ZP_00150992)	81	237	e^{-102}
				67	68	$2e^{-19}$
LESE2	362	62.4	2,4-Diacetylphloroglucinol biosynthetic protein (polyketide synthase), PhlD (<i>P. fluorescens</i>) (AAX31636)	78	120	$5e^{-45}$
LESG11	572	62.6	Transcriptional activator PltR (<i>Pseudomonas</i> sp. M18) (AAQ90169); involved in regulation of pyoluteorin synthesis	98	190	e^{-103}
LESH3	804	50.3	Transcriptional activator PltR (<i>Pseudomonas</i> sp. M18) (AAQ90169); same protein as LESG11 but different region; N-terminal sequence	97	95	$2e^{-44}$
Haemagglutinin related						
PSA	> 1000	ND	Hypothetical protein (PA2459)	98	156	$3e^{-87}$
			Hypothetical protein (<i>Bacillus thuringiensis</i>) (YP_037116)	44	74	$5e^{-10}$
			C terminus of FhaB (<i>N. meningitidis</i>) (NP_283490)	50	57	$1e^{-9}$
Extended PSA	2751	50.9	Large exoprotein COG3210 involved in haem utilization or adhesion (<i>P. aeruginosa</i> PA14) (ZP_00140193)	85	364	e^{-168}
			Variation at the C terminus; hypothetical protein (PA2462)	95	319	e^{-167}
			Variation at the C terminus; hypothetical protein (PA2459)	98	177	e^{-100}
PS9	393	49.6	Haemagglutinin-like protein (PA2462)	39	140	$3e^{-19}$
Pili related						
LESB1	> 652	58.0	Pilin biogenesis protein PilC (<i>P. aeruginosa</i>) (AAA25733)	88	127	$3e^{-58}$
			PilB (PA4526)	94	88	$3e^{-43}$
			PilC (PA4527)	89	94	$7e^{-43}$
LESE11	508	53.9	Pilin glycosylation protein PilO (<i>P. aeruginosa</i>) (CAA58769)	99	169	$1.0e^{-81}$
DNA modification/recombination/transposases						
LESC12	516	50.0	Predicted transposase (<i>Azoarcus</i> sp. EbN1) (YP_158450)	47	57	$5e^{-8}$
LESD7	458	53.9	Hypothetical protein MCA0729 (<i>Methylococcus capsulatus</i>) (YP_113236); possible helicase	82	152	$2e^{-72}$
LESE6	975	58.6	RadC, putative DNA-repair protein (<i>Vibrio cholerae</i>) (AAL59704)	73	52	$1e^{-16}$
			Periplasmic molybdate-binding protein/domain (<i>Microbulbifer degradans</i>) (ZP_00315765)	47	59	$1e^{-10}$
LESH2	> 548	59.1	Hypothetical protein BB0398 (<i>Bordetella bronchiseptica</i>) (NP_886947); putative transposase	49	101	$4e^{-20}$

Table 3. cont.

SSH sequence	Length (bp)	GC content (%)	Best BLASTX match/comments, with accession nos	ID (%)	Length (aa)	E value
LESA8	637	58.4	RadC family protein (<i>P. putida</i> KT2440) (NP_744641)	83	109	2.5e ⁻⁴³
LESB5	357	58.0	Rhs family protein (PA14) (ZP_00136740)	91	118	7e ⁻⁵⁸
			Hypothetical protein (PA2457); recombination hot spot	85	118	4e ⁻⁵³
Metabolism/enzymes						
LESD9	648	43.7	FOG: HEAT repeat (<i>Anabaena variabilis</i>) (ZP_00162995)	32	174	7e ⁻¹⁰
LESE9	727	53.4	Asparagine synthase, glutamine hydrolysing (<i>Magnetococcus</i> sp.) (ZP_00606439); N terminus	58	72	4e ⁻¹⁶
LESD10	419	60.4	Asparagine synthetase, glutamine hydrolysing (<i>Psychrobacter</i> sp.) (YP_263961)	65	126	9e ⁻⁴⁴
LESG9	422	61.1	Chitinase domain protein (<i>P. fluorescens</i> Pf-5) (YP_260898)	65	44	3e ⁻⁹
LESB2	603	57.7	Putative ATPase (<i>Nitrobacter hamburgensis</i> X14) (ZP_00627842)	20	202	8e ⁻⁵
Acyltransferases						
PS54	425	50.6	NSH			
Extended PS54	860	49.1	Acyltransferase for 30S ribosomal subunit protein S18 (<i>Yersinia pestis</i>) (NP_671044)	34	79	0.019
LESH11	556	59.9	Predicted acyltransferase (<i>Ralstonia metallidurans</i>) (ZP_00593471)	41	185	2e ⁻²³
Transport						
LESA11	638	61.9	Inner-membrane permease, component of ABC transporter system (<i>Pseudomonas</i> sp. M18) (AAV59068)	99	198	6e ⁻⁹⁴
LESF8	401	60.1	Conserved hypothetical protein (<i>Magnetococcus</i> sp.) (ZP_00608284)	28	133	2e ⁻⁵
Hypothetical proteins						
LESF5	522	59.4	Hypothetical protein (PA2790)	99	114	1e ⁻⁶¹
LESH8	>693	62.3	Hypothetical protein Paer03004107 (PA14) (ZP_00140093)	100	186	e ⁻¹⁰⁰
PS32	138	58.7	Hypothetical protein PP2496 (<i>P. putida</i> KT2440) (NP_744644)	82	45	9e ⁻¹⁸
LES5	528	60.2	Hypothetical protein PP2496 (<i>P. putida</i> KT2440) (NP_744644); matches same protein as PS32 but in different region	93	175	1.1e ⁻⁸⁴
LESC10	494	58.7	Hypothetical protein PP3887 (<i>P. putida</i> KT2440) (NP_746017)	34	150	2e ⁻¹⁵
LESD8	359	43.5	Hypothetical protein Pflu02002711 (<i>P. fluorescens</i> PfO-1) (ZP_00264577)	40	109	2e ⁻¹⁵
LESF12	765	57.7	Hypothetical protein (<i>P. fluorescens</i>) (CAE84145)	36	144	1e ⁻¹⁶
LES11	410	56.8	Hypothetical protein (<i>Pseudomonas resinovorans</i>) (NP_758702)	32	131	1e ⁻¹⁷
LESA9	463	61.6	Hypothetical protein Avin02002756 (<i>Azotobacter vinelandii</i>) (ZP_00089890)	33	145	3e ⁻⁹
LESB3	692	60.6	Hypothetical protein (PA5517)	100	73	2e ⁻³⁵
			Hypothetical protein S4804 (<i>Shigella flexneri</i>) (NP_836048)	29	84	5e ⁻⁴
LESB11	493	56.6	Hypothetical protein Daro03003346 (<i>Dechloromonas aromatica</i> RCB) (ZP_00149999)	89	91	1e ⁻⁴³
LESF1	428	57.5	Hypothetical protein Daro03003346 (<i>D. aromatica</i> RCB) (ZP_00149999); same protein as B11 but different region	92	142	2e ⁻⁷²
LESH12	1100	52.4	Hypothetical protein plu0126 (<i>Photorhabdus luminescens</i>) (NP_927496)	24	229	1e ⁻⁴
LESB6	509	62.9	Hypothetical protein DET0256 (<i>Dehalococcoides ethenogenes</i> 195) (YP_181004)	44	69	1e ⁻¹⁰
NSH						
PS1	338	42.9	NSH			
PS2	282	45.7	NSH			
PS3	250	51.6	NSH			
PS12	288	52.1	NSH			

Table 3. cont.

SSH sequence	Length (bp)	GC content (%)	Best BLASTX match/comments, with accession nos	ID (%)	Length (aa)	E value
PS15	882	47.5	NSH			
LES2	591	56.9	NSH			
LES4	>715	46.6	NSH			
LES9	639	49.8	NSH			
LESA10	480	55.8	NSH			
LESA12	360	58.3	NSH			
LESC1	571	62.2	NSH			
LESC3	483	62.9	NSH			
LESE3	477	49.5	NSH			
LESE5	419	60.4	NSH			
LESE12	735	59.9	NSH			
LESF2	233	59.7	NSH			
LESF3	816	56.3	NSH			
LESF10	549	47.7	NSH			
LESG10	981	49.1	NSH			
LESH1	400	59.5	NSH			
LESH6	657	34.6	NSH			

exception of LES4, which lies beyond the putative island. The diagnostic probe (PS21) appears to lie in a cargo region of the previously identified island PAGI-2 (Klockgether *et al.*, 2004; Larbig *et al.*, 2002) that also contains LESB9, the putative silver-efflux-related sequence (cluster 2). Twelve of the SSH sequences (in order: LESH1, LESE4, LESB4, LESD11, LESG2, LESF9, LESD7, PS32, LES5, LEA8, LESE6 and LESB2) were clustered on a different contig within a second large putative island containing a number of genes related to *C. thermocellum* (cluster 3). There were three separate contigs containing multiple SSH sequences, some of which were bacteriophage related. One contained, in order, LES1, LEA9, LESC1, LESB6, LES6 and LESG5 (cluster 4); a second contained, in order, LEA11, LESH3, LESG11, LESE2, LESG3, LEA2, LESC11, LESC10, LESH5, LEA12 and LESF3 (cluster 5); and a third contained, in order, LESC3, LESE12, LESE7, LESF8, LESF6 and LESG9 (cluster 6). There were also several other contigs containing more than one SSH sequence. Not surprisingly, one of these linked the type III pyoverdine-synthesis-related sequences LESH4 and LES8.

Distribution of subtracted sequences

The distribution of 25 LES sequences amongst a panel of strains was assessed by PCR amplification assays using oligo-nucleotide primers designed from subtracted sequences (Table 4). The sequences selected included multiple representatives for each of the six major clusters identified by preliminary analysis of the genome sequence, sequence LESD9 (which shared a contig with LESB5, PS15, LESG10 and PS2), sequence LESB3 (which shared a contig with LESH11), and the type III pyoverdine-synthesis-related sequences LESH4 and LES8, and LEA3 (which shared a contig with LEA1).

The distribution of sequences from cluster 1 confirmed that the putative genomic island was only present in LES isolates, but was completely missing from three LES isolates. LES4 is beyond the putative island and has a radically different distribution from the other SSH sequences on this contig (Table 4). All of the SSH sequences, including PS54, PS3, PS1 and PS12 from our previous study (Parsons *et al.*, 2002), were absent from LES isolates 109 (isolated in Liverpool), 8995 (isolated in Sheffield, UK) and 10333 (isolated in Nottingham, UK). The putative island is flanked by equivalents to strain PAO1 ORFs PA0831 and PA0832.

Cluster 2, including PS21 and the silver-efflux-pump-related SSH sequence LESB9, was present in all LES isolates. All four PS21 false-positive strains were also PCR positive for LESB9, confirming the linkage of these two sequences. All SSH sequences tested from cluster 3 were PCR positive for all LES isolates but PCR negative for all other genetically characterized isolates (Table 4). Further PCR assays for sequences LESE4 and LESF9 were used to screen the uncharacterized non-LES CF isolates from Liverpool and from Clatterbridge Hospital. Three isolates (CH31, CH47 and 49278) were PCR positive for LESE4, and two of these

Table 4. Distribution of subtracted sequences based on PCR amplification assays

1, LESG6; 2, LESH12; 3, LESG7; 4, LESF1; 5, LESB7; 6, LES4; 7, LESB9; 8, PS21; 9, LESF4; 10, LESD11; 11, LESF9; 12, LESE6; 13, LES1; 14, LESA9; 15, LES6; 16, LESA11; 17, LESG11; 18, LESH5; 19, LESE7; 20, LESF8; 21, LESG9; 22, LESD9; 23, LESB3; 24, LESH4; 25, LES8; 26, LESA3. N, not determined.

Isolates	Number of PCR positives for each LES subtracted sequence																																									
	Cluster 1						Cluster 2						Cluster 3						Cluster 4						Cluster 5						Cluster 6						Pyoverdine related					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26																
LES (<i>n</i> =22)	19	19	19	19	19	22	22	22	22	22	22	22	18	17	18	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22											
Midlands1 (<i>n</i> =12)	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	12	12	0	0	0	0	12	12	12	12	12	12	12	12	12	12											
Manchester (<i>n</i> =11)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11											
Non-LES, genetically different (<i>n</i> =16)	0	0	0	0	0	5	0	0	0	0	0	0	1	2	1	6	6	0	2	2	1	0	8	2	2	2	2	2	2	2	7											
Non-LES isolates, Liverpool (<i>n</i> =43)	N	N	N	N	N	N	N	0	1	N	1	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N											
Non-LES CF isolates, Clatterbridge (<i>n</i> =32)	N	N	N	N	N	N	N	0	2	N	1	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N											
PAO1 (<i>n</i> =1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0											
PA14 (<i>n</i> =1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0											
Clone C (<i>n</i> =1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0											
PS21 false positives (<i>n</i> =4)	0	N	N	N	N	N	4	4	0	N	0	0	N	0	0	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N										

isolates, CH47 and 49278, were also PCR positive for LESF9. All other PS21-negative isolates were PCR negative for each of these sequences. PFGE typing of isolates CH31, CH47 and 49278 confirmed that these did not match the LES pulsotype (data not shown).

The PCR-based distribution profiles for the three cluster 4 sequences were similar but not identical, and indicated some instability. Two LES isolates were PCR negative for all three sequences. Two of the three cluster 5 sequences (LESA11 and LESG11) shared identical distributions and were stable amongst LES and Midlands1 isolates. The third sequence tested from cluster 5 (LESH5) was greater than 50 kb away from the other two. Only LES isolates were PCR positive for this sequence. The three cluster 6 SSH sequences shared nearly identical distributions and were detected in all LES isolates. The other sequences screened were present in both LES and Midlands1 isolates. The Manchester epidemic strain isolates were PCR positive for only one of the sequences (LESA3, matching a putative siderophore receptor) tested (Table 4).

PCR-based identification of the LES

We have reported previously that the LES strain carries approximately 95 % of the ORFs present in strain PAO1 and that estimations based on PFGE indicate that the LES genome is no bigger than that of PAO1 (Salunkhe *et al.*, 2005). In our previous study (Parsons *et al.*, 2002), we identified 14 LES sequences subtracted from strain PAO1, including five (PS1, PS3, PS12, PS21 and PS54) that were LES specific. All of these, except PS21, were absent from one of the LES isolates screened (isolate 109). After considerably expanding the SSH database and conducting preliminary searches of the LESB58 genome sequence, it is clear that sequences PS1, PS3 and PS54, along with 13 other SSH sequences, are part of an LES-specific island that is associated with some instability. Thus, although we have never found this island in strains other than LES, any test for the LES based on this island may be prone to false negatives. In contrast, PS21, part of a *merA* gene located in an apparent cargo region of PAGI-2, is stable in the LES, but occurs rarely in other *P. aeruginosa* strains, making tests based on PS21 theoretically susceptible to rare false positives. Sequence LESF9 forms part of a putative genomic island present in all LES isolates but is, like PS21, occasionally also found in non-LES isolates. In the absence of a definitive LES-specific PCR assay that is positive for all LES isolates tested, we suggest a strategy using a combination of PCR assays for the detection of PS21 and LESF9. Since PS21 has proved to be an effective tool, especially in a setting where LES is widespread (Panagea *et al.*, 2003; our unpublished data), we would propose the continuation of this test. For clinicians and patients alike, it is important that all LES isolates give a positive result. In order to achieve this, in the light of the instability of the LES genome, we may have to accept the possibility of rare false positives. However, if a combination of PCR assays for PS21 and LESF9 is used, then at least with our collection of isolates,

only LES isolates are PCR positive for both of these sequences. We would suggest that any isolates that are PCR positive for just one of these sequences should be further tested using a different typing method, such as PFGE or RAPD.

The LES mobilome

The SSH data indicated that bacteriophage DNA is likely to make a significant contribution to the accessory genome of the LES. D3 is a lysogenic bacteriophage with a genome of 56.4 kb (Kropinski, 2000) (accession no. AF165214) implicated previously in serotype conversion (Newton *et al.*, 2001) and transduction (Cavenagh & Miller, 1986) in *P. aeruginosa*. SSH sequences matching widely separated regions of the D3 sequence indicate the possible presence of this bacteriophage genome within the LES. D3112 is a transposable Mu-like bacteriophage with a genome of 37.6 kb (Wang *et al.*, 2004). We identified LES sequences matching neighbouring genes within the D3112 genome, but we have no further indication concerning the extent of the D3112 DNA within the genome of the LES. Interestingly, these two sequences appear to be missing from the LES isolate chosen for genome sequencing, suggesting either that the bacteriophage has been acquired by isolate LES400 or that it has been lost by isolate LESB58. In addition, the bacteriophage-related cluster 4 was absent from some LES isolates. SSH also indicated that the LES contains a putative tail-fibre protein from another bacteriophage reported elsewhere in *P. aeruginosa* (Nakayama *et al.*, 2000). The presence of several other subtracted sequences with matches sharing lower identities with bacteriophage-related proteins, many clustered on to one of three contigs, suggests that there may be additional whole (or fragments of) bacteriophage genomes within the LES. The details will become clear when the ongoing genome sequence project is further advanced.

ACKNOWLEDGEMENTS

We would like to thank Ty Pitt and Fiona Scott for providing strains. We are also grateful to Julian Parkhill and colleagues at the Pathogen Sequencing Unit of the Wellcome Trust Sanger Unit, and Bob Hancock, Fiona Brinkman and colleagues from the *Pseudomonas* Genome Project, for their excellent work in the LES and PAO1 genome sequence projects. This work was funded by the UK Cystic Fibrosis Trust.

REFERENCES

- Al Aloul, M., Crawley, J., Winstanley, C., Hart, C. A., Ledson, M. J. & Walshaw, M. J. (2004). Increased morbidity associated with chronic infection by an epidemic *Pseudomonas aeruginosa* strain in CF patients. *Thorax* **59**, 334–336.
- Armstrong, D., Bell, S., Robinson, M. & 8 other authors (2003). Evidence for spread of a clonal strain of *Pseudomonas aeruginosa* among cystic fibrosis clinics. *J Clin Microbiol* **41**, 2266–2267.
- Bender, C., Rangaswamy, V. & Loper, J. (1999). Polyketide production by plant-associated pseudomonads. *Annu Rev Phytopathol* **37**, 175–196.
- Cavenagh, M. M. & Miller, R. V. (1986). Specialized transduction of *Pseudomonas aeruginosa* PAO by bacteriophage D3. *J Bacteriol* **165**, 448–452.
- Chambers, D., Scott, F., Bangur, R. & 7 other authors (2005). Factors associated with infection by *Pseudomonas aeruginosa* in adult cystic fibrosis. *Eur Respir J* **26**, 651–656.
- Cheng, K., Smyth, R. L., Govan, J. R., Doherty, C., Winstanley, C., Denning, N., Heaf, D. P., van Saene, H. & Hart, C. A. (1996). Spread of beta-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic. *Lancet* **348**, 639–642.
- Choi, J. Y., Sifri, C. D., Goumnerov, B. C., Rahme, L. G., Ausubel, F. M. & Calderwood, S. B. (2002). Identification of virulence genes in a pathogenic strain of *Pseudomonas aeruginosa* by representational difference analysis. *J Bacteriol* **184**, 952–961.
- Croft, L., Beatson, S. A., Whitchurch, C. B., Huang, B., Blakeley, R. L. & Mattick, J. S. (2000). An interactive web-based *Pseudomonas aeruginosa* genome database: discovery of new genes, pathways and structures. *Microbiology* **146**, 2351–2364.
- Edenborough, F. P., Stone, H. R., Kelly, S. J., Zadik, P., Doherty, C. J. & Govan, J. R. (2004). Genotyping of *Pseudomonas aeruginosa* in cystic fibrosis suggests need for segregation. *J Cyst Fibros* **3**, 37–44.
- Ernst, R. K., D'Argenio, D. A., Ichikawa, J. K. & 12 other authors (2003). Genome mosaicism is conserved but not unique in *Pseudomonas aeruginosa* isolates from the airways of young children with cystic fibrosis. *Environ Microbiol* **5**, 1341–1349.
- Espinosa-Urgel, M., Salido, A. & Ramos, J. L. (2000). Genetic analysis of functions involved in adhesion of *Pseudomonas putida* to seeds. *J Bacteriol* **182**, 2363–2369.
- Jones, A. M., Govan, J. R., Doherty, C. J., Dodd, M. E., Isalska, B. J., Stanbridge, T. N. & Webb, A. K. (2001). Spread of a multiresistant strain of *Pseudomonas aeruginosa* in an adult cystic fibrosis clinic. *Lancet* **358**, 557–558.
- Klee, S. R., Nassif, X., Kusecek, B., Merker, P., Beretti, J. L., Achtman, M. & Tinsley, C. R. (2000). Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae*. *Infect Immun* **68**, 2082–2095.
- Klockgether, J., Reva, O., Larbig, K. & Tümmler, B. (2004). Sequence analysis of the mobile genome island pKLC102 of *Pseudomonas aeruginosa* C. *J Bacteriol* **186**, 518–534.
- Kropinski, A. M. (2000). Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J Bacteriol* **182**, 6066–6074.
- Kus, J. V., Tullis, E., Cvitkovitch, D. G. & Burrows, L. L. (2004). Significant differences in type IV pilin allele distribution among *Pseudomonas aeruginosa* isolates from cystic fibrosis (CF) versus non-CF patients. *Microbiology* **150**, 1315–1326.
- Larbig, K. D., Christmann, A., Johann, A., Klockgether, J., Hartsch, T., Merkl, R., Wiehlmann, L., Fritz, H. J. & Tümmler, B. (2002). Gene islands integrated into tRNA^{Gly} genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J Bacteriol* **184**, 6665–6680.
- Lewis, D. A., Jones, A., Parkhill, J., Speert, D. P., Govan, J. R., Lipuma, J. J., Lory, S., Webb, A. K. & Mahenthalingam, E. (2005). Identification of DNA markers for a transmissible *Pseudomonas aeruginosa* cystic fibrosis strain. *Am J Respir Cell Mol Biol* **33**, 56–64.
- Liang, X., Pham, X. Q., Olson, M. V. & Lory, S. (2001). Identification of a genomic island present in the majority of pathogenic isolates of *Pseudomonas aeruginosa*. *J Bacteriol* **183**, 843–853.
- Locht, C., Bertin, P., Menozzi, F. D. & Renauld, G. (1993). The filamentous haemagglutinin, a multifaceted adhesion produced by virulent *Bordetella* spp. *Mol Microbiol* **9**, 653–660.

- McCallum, S. J., Corkill, J., Gallagher, M., Ledson, M. J., Hart, C. A. & Walshaw, M. J. (2001). Superinfection with a transmissible strain of *Pseudomonas aeruginosa* in adults with cystic fibrosis chronically colonised by *P. aeruginosa*. *Lancet* **358**, 558–560.
- McCallum, S. J., Gallagher, M. J., Corkill, J. E., Hart, C. A., Ledson, M. J. & Walshaw, M. J. (2002). Spread of an epidemic *Pseudomonas aeruginosa* strain from a patient with cystic fibrosis (CF) to non-CF relatives. *Thorax* **57**, 559–560.
- Nakayama, K., Takashima, K., Ishihara, H. & 7 other authors (2000). The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol Microbiol* **38**, 213–231.
- Nelson, K. E., Weinel, C., Paulsen, I. T. & 40 other authors (2002). Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* **4**, 799–808.
- Newton, G. J., Daniels, C., Burrows, L. L., Kropinski, A. M., Clarke, A. J. & Lam, J. S. (2001). Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. *Mol Microbiol* **39**, 1237–1247.
- O'Carroll, M. R., Symmis, M. W., Wainwright, C. E., Greer, R. M., Mitchell, P., Coulter, C., Sloots, T. P., Nissen, M. D. & Bell, S. C. (2004). Clonal strains of *Pseudomonas aeruginosa* in paediatric and adult cystic fibrosis units. *Eur Respir J* **24**, 101–106.
- Panagea, S., Winstanley, C., Parsons, Y. N., Walshaw, M. J., Ledson, M. J. & Hart, C. A. (2003). PCR-based detection of a cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*. *Mol Diagn* **7**, 195–200.
- Parkhill, J., Achtman, M., James, K. D. & 25 other authors (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502–506.
- Parsons, Y. N., Panagea, S., Smart, C. H. M., Walshaw, M. J., Hart, C. A. & Winstanley, C. (2002). Use of subtractive hybridization to identify a diagnostic probe for a cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*. *J Clin Microbiol* **40**, 4607–4611.
- Potvin, E., Lehoux, D. E., Kukavica-Ibrulj, I., Richard, K. L., Sanschagrín, F., Lau, G. W. & Levesque, R. C. (2003). In vivo functional genomics of *Pseudomonas aeruginosa* for high-throughput screening of new virulence factors and antibacterial targets. *Environ Microbiol* **5**, 1294–1308.
- Rahme, L. G., Stevens, E. J., Wolfort, S. F., Shao, J., Tompkins, R. G. & Ausubel, F. M. (1995). Common virulence factors for bacterial pathogenicity in plants and animals. *Science* **268**, 1899–1902.
- Römling, U., Wingender, J., Müller, H. & Tümmler, B. (1994). A major *Pseudomonas aeruginosa* clone common to patients and aquatic habitats. *Appl Environ Microbiol* **60**, 1734–1738.
- Salunkhe, P., Smart, C. H. M., Morgan, J. A. W., Panagea, S., Walshaw, M. J., Hart, C. A., Geffers, R., Tümmler, B. & Winstanley, C. (2005). A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance. *J Bacteriol* **187**, 4908–4920.
- Scott, F. W. & Pitt, T. L. (2004). Identification and characterization of transmissible *Pseudomonas aeruginosa* strains in cystic fibrosis patients in England and Wales. *J Med Microbiol* **53**, 609–615.
- Smith, E. E., Sims, E. H., Spencer, D. H., Kaul, R. & Olson, M. V. (2005). Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. *J Bacteriol* **187**, 2138–2147.
- Spencer, D. H., Kas, A., Smith, E. E., Raymond, C. K., Sims, E. H., Hastings, M., Burns, J. L., Kaul, R. & Olson, M. V. (2003). Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*. *J Bacteriol* **185**, 1316–1325.
- Stover, C. K., Pham, X. Q., Erwin, A. L. & 28 other authors (2000). Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**, 959–964.
- Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. & Swaminathan, B. (1995). Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33**, 2233–2239.
- Wang, P. W., Chu, L. & Guttman, D. S. (2004). Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. *J Bacteriol* **186**, 400–410.
- Weinel, C., Nelson, K. E. & Tümmler, B. (2002). Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environ Microbiol* **4**, 809–818.
- Winstanley, C. (2002). Spot the difference: applications of subtractive hybridization to the study of bacterial pathogens. *J Med Microbiol* **51**, 459–467.
- Winstanley, C. & Hart, C. A. (2000). Presence of type III secretion genes in *Burkholderia pseudomallei* correlates with Ara[−] phenotypes. *J Clin Microbiol* **38**, 883–885.
- Wolfgang, M. C., Kulasekara, B. R., Liang, X., Boyd, D., Wu, K., Yang, Q., Miyada, C. G. & Lory, S. (2003). Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **100**, 8484–8489.